Fault Tolerance in the Orlando Full Service Network

Author: **Michael Adams**, Time Warner Cable Advanced Engineering
(Email: **michael.adams@twcable.com**)

Abstract

This paper discusses the notion of fault tolerance in an interactive multi-media delivery system. The various components are reviewed together with strategies for hardware redundancy and the software mechanisms necessary to support the use of hardware redundancy. Traditional telecommunications approaches to fault tolerance are referenced and the cost and complexity of these schemes is shown to be unrealistic for the kinds of full service networks envisioned. The Orlando Full Service Network is described and its expected and actual failure modes are discussed. Fault tolerance mechanisms that are designed to use spare capacity in the delivery system are described and the effectiveness and simplicity of alternative schemes are discussed. The paper proposes a set of network design rules that can be used to build fault-tolerant entertainment-delivery networks without increasing cost dramatically.

Introduction

The Orlando Full Service Network is designed to provide interactive television services (including movies-on-demand, home-shopping and video games) to a community of 4,000 subscribers. In a system of this size, it is important to understand the possible failure modes and their impact on system availability. To meet service availability metrics, it is important that certain faults can be tolerated by the system to eliminate or reduce down-time due to failures. Thus we have a requirement for a Fault Tolerant system design.

System and Server Availability Requirements

When designing any system certain requirements for system and service availability can be defined. There are a number of ways of doing this but two of the most common metrics are MTBF and MTTR.

Mean Time Between Failure (MTBF) is a way of measuring how often (on average) a component or the entire system can fail. Typically, in a product development environment, the MTBF of each component is modeled and the expected MTBF of the entire system can be predicted statistically.

Mean Time To Repair (MTTR) is another important metric, because after a failure has occurred, the key to returning the system to service is the time (on average) it takes to repair the system.

Using MTBF and MTTR metrics allows the designer to predict the expected availability of the system and of the service that the system provides. If certain system failure modes only affect part of the system, then the service availability will be different from the system availability. For example, in the FSN a failure of a single server may cause a loss of service to a fraction of the subscribers, but not impact the other subscribers at all.

Resource Management Requirements

In a complex network like the Orlando FSN, there are many failures that can and will occur. These may be as simple as a loose connector, or as complex as a software bug in a Media Server. In the event of such a failure, a well designed system will respond in two ways:

- The system will attempt to use other resources to replace the failed resource. For example, if a DS3 link fails, the traffic for that link may be re-routed over other DS3 links (assuming there is sufficient spare capacity).

- The system will generate an alarm to notify a technician of the failure. The failed component must be isolated and replaced as quickly as possible to restore the system to full operation. In a complex system this fault isolation step can be very complex to do manually. A network management system can reduce the time to isolate a fault by using a rule-based diagnostic tool. A set of rules is developed over time by the system designers and technicians. The diagnostic tool uses these rules to infer which component has failed and produces a report for the technician on duty.

System Overview

A system overview diagram is presented below. The key elements that are monitored are the Media Servers, the ATM Switch, the ATM Multiplexers (MUX) and the Home Communications Terminals (HCT).

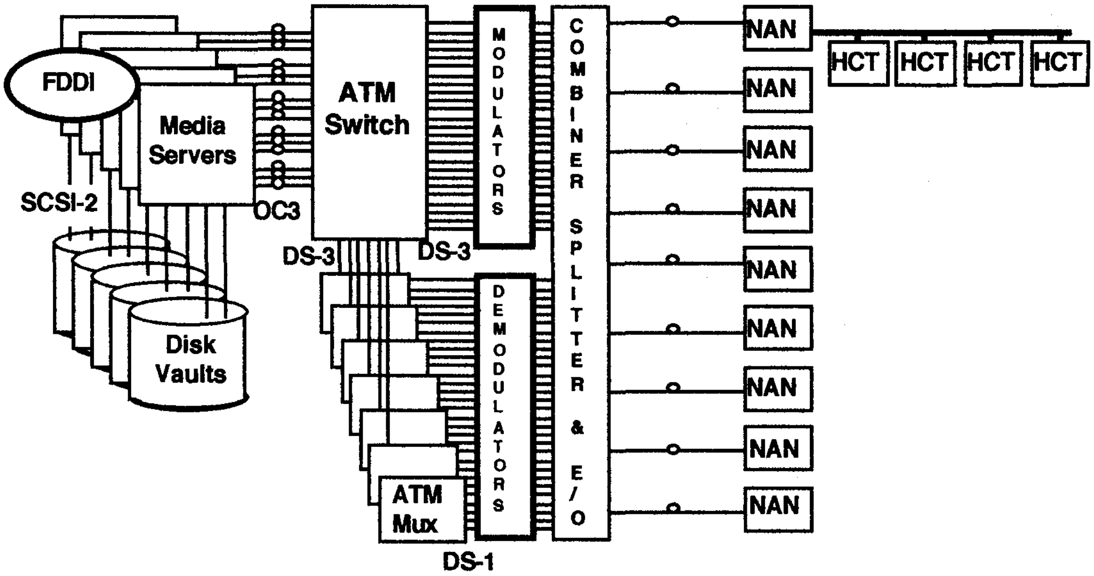

Figure 1. System Overview

The major system components are shown in Figure 1. The impact on system availability of each component is described below.

- The ATM Switch is a single device, which provides all connectivity from the Media Servers to the HCTs. If the ATM switch were to fail completely, interactive services would be interrupted to **all** FSN subscribers. (However, analog cable services would still be available).

- There are 7 ATM Multiplexers, which operate independently of each other. Therefore the complete failure of a multiplexer would interrupt interactive services to approximately one seventh of the FSN subscribers. (However, analog cable services would still be available).

- There are 152 QAM modulators. Each serves up to 7 active, interactive subscribers.

- There are 16 QPSK modulators. Each serves an entire neighborhood (providing control and timing information to all HCTs in that neighborhood).

- There are 240 QPSK demodulators. 160 modulators are dedicated to reverse signaling from interactive subscribers and each is shared between 25 subscribers. (The remaining 80 modulators are available to provide services requiring higher bandwidth return, such as video telephony or data services).

All above equipment is located in the Network Operations Center (NOC) and is powered by a conditioned, un-interruptable power supply. Reserve power is supplied by batteries and a backup diesel generator. The headend equipment is housed in approximately 1000 square feet of computer room which is air-conditioned and equipped with a raised floor. Fire protection is provided by a Halon gas extinguishing system.

There are a total of approximately 500 physical interfaces between the various components in the NOC, and thousands of optical and electrical connectors and cables. The ability to quickly isolate and repair any problems as they occur is very important in this complex a system.

- There are 16 neighborhoods (of approximately 250 subscribers each). Each neighborhood is served by a laser transmitter and receiver at the headend and Neighborhood Area Node (NAN) in the field. The NAN provides optical to electrical conversion at the point where the fiber ends and the coax network begins. The failure of a headend transceiver, the fiber or the NAN would interrupt interactive and analog services to all subscribers in a neighborhood. The NAN is powered by conventional, battery-backed power-supplies via the coaxial plant.

- There are 4000 Home Communications Terminals, each located in a subscriber residence. Obviously, the failure of an HCT impacts only a single subscriber, however it is more difficult to service because of its location. The HCT repair strategy is unit-replacement. It is very important that failures in the system that manifest themselves as an HCT not working do not result in a wasted visit to the subscriber.

This is a key difference between a traditional cable system which, being completely broadcast in nature, does not usually experience these kinds of failure. In the Full Service Network, which is much more transactional in nature, it is often not obvious whether a failure is in the HCT or some component in the NOC.

System Failure Modes

The preceding discussion describes some of the ways in which the various components that make up the system can fail. It is very important to understand these failure modes because the failure of only part of the system is not necessarily catastrophic as long as:

- **The failure does not affect the rest of the system.** For example, the failure of an HCT would seem to be a trivial case, but a failure mode in which the HCT's reverse transmitter starts to send continuously would impact the other 24 HCTs sharing that reverse QPSK channel. (In this example, the solution is to implement a reset command to allow the system to disable the offending HCT and localize the failure to that subscriber).

- **The failed component can be replaced without impact to the rest of the system.** This may sound obvious, but it is remarkable how certain failure modes seem to appear which require a complete system outage to clear them. These failure modes have a major impact on system availability. It is important to recognize these failure modes in the design phase so that they can be 'designed out' of the system.

For each component that can fail, a decision has to be made: **Is it worthwhile to design-in fault tolerance?** The design philosophy for each major component is discussed below.

Media Server

The FSN uses eight Silicon Graphics Challenge XL servers. The server hardware is not fault-tolerant. (The cost of providing hardware fault tolerance would more than double the cost of the media servers). However, the media servers form a fault-tolerant distributed computer system. At a software level, redundancy is provided in each server to spare the others. The design of such a system is complex, however not all functions need to be duplicated. Special design attention to redundancy was given to following critical components

Connection Manager

The connection manager is an example of a critical function that should not fail if a server fails. The connection manager allocates an ATM Virtual Channel in response to an application request. To make the connection manager reliable even when a server fails, it is split into Neighborhood Connection Managers (NCM). Each NCM is duplicated in software such that the backup NCM runs on a different server than the active NCM.

If the active NCM fails (because the server hardware fails), the backup NCM can take-over without loss of service. When the server recovers (usually after a re-start), the active NCM is restarted and resumes its normal operation. To allow the changeover to be seamless, the active and backup NCMs checkpoint the current connection information to a shared file.

Software Processes

Any software process may fail due to a hardware or software exception. The general strategy is to restart any process that fails to recover from exceptions. Typically each server has hundreds of software processes running at any moment in time, and the failure of a single process should not affect the integrity of the others. This is achieved by using Virtual Memory techniques that provide each process with its own unique address space. Even if the process attempts to write to random memory addresses, the memory management hardware will protect the other processes from being affected.

Disk Vaults

Each server is connected to a large number of disk drives which are housed in disk vaults. The drives are connected using a SCSI-2 interface controller for each 16 disk drives. This approach reduces the impact of the failure of a single disk drive or SCSI-2 controller.

A number of more sophisticated approaches exist to provide for fault tolerance. These are generally known as RAID - Redundant Array of Inexpensive Disks. There are actually five main flavors of RAID from RAID-1 through RAID-5. Some approaches are reliable but store two copies of the data (mirroring), and so require twice the disk storage. Other approaches achieve reliability at lower cost by storing additional parity information that can be used to reconstruct data lost due to a disk failure.

ATM Switch

The FSN uses a single AT&T GCNS-2000 ATM Switch. As previously noted, the ATM switch is a single point of failure for the entire system. This dictated the selection of a highly reliable switch[1]. Fortunately this is a common requirement in the telecommunications industry and the development of extremely reliable switching platforms has evolved to meet this requirement.

AT&T are experienced in the design of highly reliable switches. (The Bellcore recommendation for ATM switches is a maximum of 2 hours downtime in 40 years!). The AT&T GCSN-2000 design philosophy eliminates all single points of failure. This is achieved by duplicating all critical components. Each active

[1]. Note that the chosen solution is not the only one available. An alternative is to implement a reliable network based on multiple simplex switches instead of a single duplex switch. Network redundancy (as this is called) has the advantage of using less expensive, simplex switches and of enabling the use of '1 for n' sparing. For example, 10 simplex switches could be spared by an eleventh switch. A further extension is to build a system with 10 switches and accept the reduced impact of a single switch failure.

component is monitored and in the event of a failure, the backup component takes over the tasks of the failed component without any interruption.

Another important requirement is that it must be possible to repair the switch while it continues to operate normally. (This is analogous to replacing a tire on a car while driving it down a freeway at 55 miles per hour). In order to allow this, it must be possible to 'hot-swap' cards while the system is powered. This requires special power-up circuitry and connectors on the cards and card-cages.

In the ATM switch the following components are duplicated:

Control Processor

The control processor is responsible for monitoring the status of the switch and allowing the provisioning of ATM connections. The entire control processor is duplicated and each processor monitors the other. Each is capable of running the entire switch if the other fails.

Shelf Controllers

The Shelf Controller is responsible for monitoring the status of a shelf and allowing the provisioning of ATM connections. The shelf controller is duplicated and each shelf controller monitors the other. The links to the control processor are also duplicated.

Fabric Interface

The interface from the shelf to the fabric is duplicated. However, each fabric interface is only connected to one of the two switching fabrics. This means that the system can not tolerate two fabric interfaces on two different shelves. This is a double failure and is extremely unlikely in a given time window.

ATM Switch Fabric

The ATM Switch fabric is duplicated and each fabric is capable of supporting the full traffic capacity of the switch. In the event of a failure of either fabric, the other fabric is able to take over with the loss of only a small amount of traffic. In practice, we found the traffic loss would not normally be noticeable to a subscriber viewing a movie.

Clock Distribution Cards

The Clock Distribution Cards are duplicated and monitored by the shelf controllers. If the active card fails, the standby card takes over.

Line Interface Cards

In addition, the Line Interface Cards can be duplicated, but in our application this was not required because the loss of a single line card does not affect a large number of subscribers. (Approximately 100 subscribers would be impacted (worst-case) by the failure of an OC3c line card).

The physical layout and location of the components is illustrated in the diagram below. Note that the physical dimensions of the switch are significantly larger than for an equivalent simplex switch.

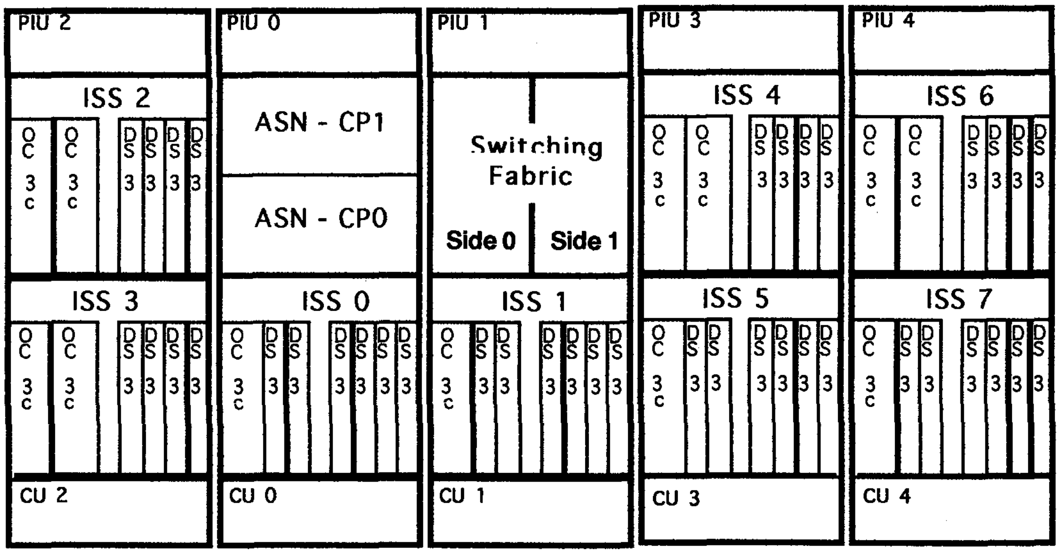

Figure 2. ATM Switch Physical Layout

ATM Multiplexer

The FSN uses seven Hitachi AMS-5000 ATM multiplexers. The multiplexer is responsible for the forward control and reverse application channels to 2 or 3 neighborhoods (400-600 subscribers). As such, it must be reliable and should not be subject to a single point of failure. The multiplexer has the following redundant components:

Control Processor

The control processor consists of 2 cards and can be duplicated. In our application this was not considered necessary because the multiplexer will continue to handle traffic even if the processor fails. (Only the provisioning and monitoring functions are affected).

Backplane Interconnect

The traffic is switched from one interface card to another by being routed over a shared-bus backplan. Each interface card acts as a repeater in the bus. To prevent the failure (or removal) of a card from interrupting the traffic, the bus can dynamically re-configure around a failure.

Interface Cards

The interface cards are not duplicated in this system. The failure of a single DS1 transmitter card could affect interactive services on up to 3 neighborhoods. New options for making this card redundant are being considered by the Orlando personnel.

The failure of a single DS1 receiver card impacts approximately 100 subscribers and is therefore not considered critical.

Modulators and Demodulators

All of the RF modulators and demodulator are supplied by Scientific Atlanta.

Each QAM modulator serves up to 7 active subscribers at any point in time. If a modulator fails, only 7 subscribers would be affected. An exception to this is the modulator used to broadcast the kernel software to the HCT. This is a single point of failure in the system but it is not duplicated because it only affects the ability of HCTs to boot.

Each QPSK modulator serves an entire neighborhood (about 200-300 subscribers). The QPSK modulator is not duplicated.

Each QPSK demodulator is shared between about 25 subscribers. The QPSK demodulator is not duplicated.

The Laser Transmitters

The laser transmitters and receivers are supplied by Scientific Atlanta.

There is a single Laser Transmitter for each neighborhood. The optical signal may be split 2 or 3 ways to the physical nodes in the plant. The Laser Transmitter is not redundant. If it should fail it would typically affect about 200 to 300 subscribers.

The Neighborhood

The neighborhood is not a fault-tolerant unit. This follows the tradition cable system view for upgraded plant. The components from the laser transmitter in the head-end to the NAN are all of low complexity, and the transmit path is analog. To duplicate the equipment and fiber, and to provide protect-switching (as SONET does) would increase the complexity and cost to an unacceptable degree.

The Home Communications Terminal (HCT)

The HCT has no requirement for fault-tolerance as it provides service only to a single subscriber. However, it must be reliable so that it does not significantly reduce the service availability to the subscriber. It must also not affect any other subscribers if it fails. This requirement is partly satisfied by the coaxial network itself which includes resistive power taps that prevent event a short-circuit at the drop from affecting other subscribers. However, if the HCT starts to transmit continuously it would impact the other HCTs sharing the reverse QPSK channel. In this case, all HCTs sharing the reverse channel will be effectively jammed, and the regular heart-beat exchange between the Server and the HCT will fail. This failure will be reported to the Network Management System.

Network Management

The inclusion of a sophisticated Network Management System (NMS) in the FSN does not affect the MTBF of any service-related component. However the NMS helps to reduce the MTTR when a failure occurs. The NMS does this in two ways:

- By monitoring the system components, an alarm indication can be generated as soon as a failure occurs.
- By building a set of rules into the NMS, the NMS can greatly assist in Fault Isolation. This reduces MTTR by helping focus repair activity on the correct component. In a complex system it can be extremely difficult to locate the failed component without this kind of assistance.

Power Supply and Air Conditioning

The physical and electrical environment in which the system operates is often overlooked. The Orlando Network Operations Center was designed to meet Telco standards which include reliable systems for:

- Un-interruptable Power Supplies - this includes battery power and a backup generator. In a normal telephony operation, all systems are required to run from a - 48 volt supply. In the FSN, only the ATM equipment meets this requirement and so invertors are required to support the Media Servers, Modulators and Demodulators and AM Laser equipment.
- Air Conditioning - this must also be supplied with backup power or temperature rise would cause equipment to automatically power down after a number of hours (depending on ambient temperature).

Conclusions

The system described is now in commercial service in Orlando, Florida. The metrics of service availability, MTBF and MTTR are being tracked for the various components. This information will be extremely valuable in the design of future networks. Designing fault tolerance into a system is only justified when failure rates and loss of service impacts are greater than a certain thresh-hold. This thresh-hold also changes with the set of applications and the subscriber's expectations, which are both in a state of flux.

Future applications (such as Telephony and Data Communications), have different service availability requirements from interactive television. The experience gained in Orlando will be especially valuable in constructing future networks to deliver a wide range of communications and entertainment services.